19th INTERNATIONAL CONGRESS ON ACOUSTICS
MADRID, 2-7 SEPTEMBER 2007

TIMBRE CONTROL OF A REAL-TIME PERCUSSIVE SYNTHESIZER

PACS: 43.66.Jh

ARAMAKI, Mitsuko[1 2] ; KRONLAND-MARTINET, Richard[2] ; VOINIER, Thierry[2] ; YSTAD, Solvi[2]
[1] CNRS-Institut de Neurosciences Cognitives de la Méditerranée, 31, chemin Joseph Aiguier 13402 Marseille Cedex 20 France; aramaki@lma.cnrs-mrs.fr
[2] CNRS-Laboratoire de Mécanique et d'Acoustique, 31, chemin Joseph Aiguier 13402 Marseille Cedex 20 France; {kronland,voinier,ystad}@lma.cnrs-mrs.fr

ABSTRACT

In this paper, we present the design and the control of a real-time synthesis model dedicated to percussive sounds. We mainly address the simulation of the "sound colorations" (i.e. metallic, wooden or crystal-clear sounds) independently of the mechanical properties of the sound source and their control using a few perceptually relevant descriptors. The synthesis model takes into account both physical and perceptual considerations, leading to a combination of additive and subtractive synthesis processes. The manipulation of the synthesis parameters is not intuitive since the relationship between the parameters and the resulting sound is intricate, especially when dealing with material control. For that, we propose a timbre control space based on the results obtained from listening tests aiming at better understanding how the perception of material in terms of sound colorations can be linked to the synthesis parameters. This timbre control space is mainly based on damping and roughness and allows for an intuitive navigation across different material categories.

INTRODUCTION

Several studies aiming at defining a timbre control space, allowing for the mapping between synthesis parameters and the perception of sound categories [16], have led to the development of the synthesizer presented here. A synthesis model of percussive sounds that simulates the solutions of the mechanical equations describing the object vibrations under free oscillations has been used [2]. In order to develop a real-time interface that offers an easy and intuitive control of the parameters, perceptually relevant signal features that listeners use as clues to distinguish different types of materials have to be identified. The parameters of interest can be considered as invariants in the sense that they are responsible for the recognition of the material, independently of the excitation, the size or the shape of the sound producing object. The identification of invariants is useful, since it provides a direct relationship between sound perception and the synthesis parameters and consequently, allows for an easy manipulation of intrinsic sound properties (e.g., add a metallic aspect to a non-metallic sound, morph different materials or different musical instruments . . .). Also, within the field of virtual reality, it is important to be able to construct sounds from a semantic description in order to adapt a sound to a visual scene.

Regarding the perception of impact sounds, previous acoustic studies have highlighted the relationships between acoustic descriptors and perception of the physical characteristics of the sound source such as the nature of the action [6], of the object [14, 9, 12, 5] and of the material [17, 8, 12]. So far, we have focused on this last aspect to construct a perceptually calibrated timbre control space for impacted materials. From the litterature, the perception of material is mainly correlated with the damping of spectral components. However, damping alone does not allow to define the timbre of impacted materials, such as Wood, Metal and Glass, because these sounds generally differ both by their time and frequency energy distributions. For instance, results of digital sound synthesis studies have shown that it is possible to create sounds that evoke different materials, although their damping characteristics are identical (sound examples available at http://www.lma.cnrs-mrs.fr/∼kronland/JASA_Categorization/sounds.html). Hence, we examined the relevance of damping together with several acoustic descriptors known to be important for timbre perception (i.e. spectral centroid, Spectral Bandwidth, Spectral Flux, Roughness) through a sound categorization test. Three different material categories (i.e. wood, metal and glass) were studied. Behavioral data were collected and changes in brain electrical activity (Event Related Potentials) were recorded to investigate the temporal dynamics of the brain processes involved in the perception and categorization of these sounds. The results from these tests were further used to define a material space for an intuitive control of the percussive synthesizer. In this paper we first briefly describe the synthesis model and its real-time implementation. Then we describe how the acoustic parameters responsible for the differentiation of material classes have been identified. Finally, a detailed description of the control of the timbre space is given.

SOUND SYNTHESIS MODEL OF PERCUSSIVE SYNTHESIZER

To address the control issue of impacted sounds, we designed a sound synthesis model taking into account both physical and perceptual aspects related to sounds (a detailed description of the model can be found in [2]). The model we propose is an extension of that proposed by Smith and Van Duyne, developed to simulate the soundboard's influence on piano tones [13]. This model is based on a time-varying subtractive synthesis process that acts on a noisy input signal. This sound-synthesis model reproduces main contributions characterizing the perceived material (determined by damping) and the perceived size and shape of the impacted object (determined by pitch and modal density). We decided to consider these two contributions separately, even if they cannot be totally disconnected from a physical point of view. In practice, the model is decomposed into three main elements:

- The material element: Material is strongly characterized by the damping, which is frequency-dependent. High frequency modes are generally more heavily damped than low-frequency modes. Actually, the dissipation of vibrating energy owing to the coupling between the structure and the air increases with frequency (see for example, [4]). To take into account this fundamental sound behavior from a synthesis point of view, a time varying filtering technique has been chosen.
- The object element: It is well known that the size and the shape of an object's attributes are mainly perceived by the pitch of the generated sound and its spectral richness. The perception of the pitch primarily correlates with the vibrating modes [5]. For complex structures, the modal density generally increases with the frequency, so that high frequency modes overlap and become indiscernible. To provide a subjective notion of the size and shape of the sounding object, a few spectral components are added to a white noise simulating the high density of modes. From a physical point of view, these spectral components mainly correspond to the eigen modes of the structures. These modes can be deduced for simple cases from the movement equation and can be simply generated by using sinusoids
- Excitation element: To model the excitation, a band-pass filter is used to control the bandwidth of the generated spectrum. From a physical point of view, the response of this filter is strongly related to the strength of the impact, that is, the bandwidth increases as a function of the impact velocity.

This model has been implemented with Max/MSP allowing a control process in real-time. The input signal of the model consists of a stochastic contribution (limited here to a Gaussian noise generator) providing a broadband spectrum and a tonal contribution simulating the emergent modes. The material element is simulated by 24 damping coefficients allowing the control for the evolution of the spectrum through 24 frequency Bark bands, corresponding to the critical bands of hearing [18]. This configuration allows the reproduction of the frequency dependency of damping, where the damping coefficients are assumed to be constant in each Bark band. An important aspect of the model is its ability to resynthesize natural sounds, meaning that one can also reproduce a given impact sound that is perceptually identical to the original. We refer the reader to a more detailed article [2] for information on this inverse problem.

TIMBRE CONTROL SPACE FOR SOUNDS FROM IMPACTED MATERIALS

Actually, the current version of the synthesizer is controlled by more than hundred parameters and the piloting aspect is still an open problem. We focused here on the control of the material “element". Hence, we studied the categorization of sounds from different impacted materials to determine the acoustic parameters perceptually relevant for an ecological synthesis of these sounds (see [1] for more details). A sound categorization test was designed based on the three following material categories: Wood, Metal and Glass. We constructed a set of stimuli by recording, analyzing and resynthesizing sounds from everyday life objects made of these materials (i.e. impacted wooden beams, metallic plates and various glass recipients). The pitches of the resynthesized sounds are different since they correspond to sounds produced by various impacted objects. To minimize influences induced by pitch changes, sounds were tuned to the same fundamental pitch (note C) but differed by 1, 2 or 3 octaves depending upon the material. In particular, since Glass sounds are characterized by higher pitches than Metal or Wood sounds, they can not be transposed to lower pitches without loosing their specificities.

Then, we constructed 15 different sound continua, in particular, 5 continua for each material transition (i.e. {Wood↔Metal}, {Wood↔Glass} and {Glass↔Metal}). These continua simulate a continuous transition between sounds of 2 different materials. We used a morphing process based upon an interpolation between the values of the parameters corresponding to the amplitudes and frequencies of spectral components of the extremes sounds. Regarding damping, we chose to apply the morphing on damping parameters {α_G,α_R} (which are defined below) instead of on the damping coefficient of each spectral component. For that, we modeled the frequency-dependency of damping $\alpha(\omega)$ by fitting the damping coefficients by an exponential function characterized by 2 parameters:

$$\alpha(\omega) = e^{(\alpha_R \omega + \alpha_G)} \qquad \text{(Eq. 1)}$$

We defined {α_G,α_R} as damping parameters. From a perceptual point of view, the parameter α_G expresses a global damping highly linked to the sound duration. The parameter α_R expresses a relative damping and quantifies the differences between damping of high frequency components and damping of low frequency components.

During the sound categorization task, the participants were asked to classify sounds of all continua presented pseudo-randomly as Wood, Metal or Glass. We collected the participants' responses and Reaction Times (RTs) for each sound. Based on these behavioral data, we determined a set of typical sounds (located at the extreme positions of the continua) of each material category by grouping sounds which were categorized as such by more than 70 % of the participants.

We then conducted acoustic analysis to examine correlations between sound categories as revealed by the behavioral data and specific acoustic parameters. In particular, we examined the damping parameters {α_G,α_R} and several acoustic descriptors known to be relevant for timbre perception [11]: spectral flux (SF), spectral centroid (SCG), spectral bandwidth (SB) and roughness (R). The attack time was not considered since it was identical for all our stimuli.

Regarding the calculation of roughness, we chose the model proposed by Vassilakis who found better correlations with roughness ratings obtained in his set of perceptual experiments than predictions of previous models (see arguments in [15]). The calculation consists in summing up the partial roughness r_{ij} for all pairs of frequency components (i,j) contained in the sound:

$$r_{ij} = 0.5 \times (a_i a_j)^{0.1} \times \left(\frac{2\min(a_i,a_j)}{a_i + a_j} \right)^{3.11} \times \left(e^{-3.5s|f_i - f_j|} - e^{-5.75s|f_i - f_j|} \right) \qquad \text{(Eq. 2)}$$

where a_i and a_j, f_i and f_j are respectively the amplitudes and frequencies of a pair of components (i,j) and the parameter s is defined by:

$$s = \frac{0.24}{0.0207 \times \min(f_i,f_j) + 18.96} \qquad \text{(Eq. 3)}$$

Inline with several studies, the acoustic analysis revealed the importance of damping (characterized by α_G and α_R) to explain the categorization of materials. As expected, we observed that damping was not sufficient, in particular, to clearly discriminate Metal from Glass, as shown in Figure 1(a). This result is in acordance with the conclusions of Giordano et al. who found that sound perception within the macro-category Metal-Glass is also influenced by the spectral properties of sounds [7]. Thus, we added the R dimension (the most relevant candidate among the timbre descriptors to discriminate Metal from both Wood and Glass) to the damping to improve the distinction between Metal and Glass. Consequently, we defined a three-dimensional space dedicated to the control of timbre across different material categories.

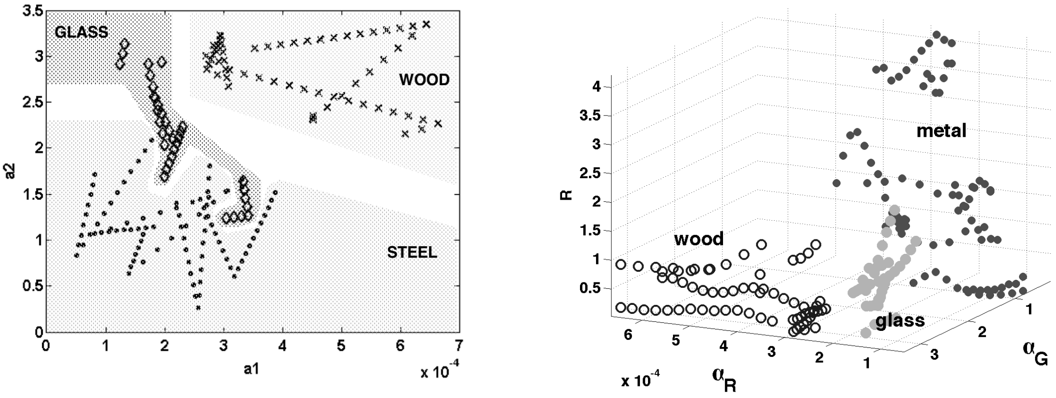

Figure 1: **(a)** Calibration of the damping parameters α_G and α_R, from the typical sounds represented for each material: Wood ($\times$), Glass ($\diamondsuit$) and Metal ($\bullet$). Hence, a specific domain was determined for each material category. **(b)** Position of typical sounds (empty circles for Wood, grey circles for Glass and black circles for Metal) in a 3D space defined by the 2 damping parameters $\{\alpha_G,\alpha_R\}$ and the roughness R.

CONTROL FOR THE TIMBRE DESCRIPTORS

Damping parameters

The material element of the synthesizer is controlled by 24 values of damping coefficients (one value per Bark band). The previous results let us propose a specific preset, by reducing the damping control to only 2 damping parameters α_R and α_G. The behavioral data allowed to calibrate the damping parameters α_G and α_R and consequently, to propose a bi-dimensional control space of materials by means of a global damping and a relative damping. Nevertheless, as illustrated in Figure 1(a), the differentiation between Metal and Glass based on $\{\alpha_G,\alpha_R\}$ is not sufficient. Therefore, an additional dimension is needed to improve the navigation between materials.

Roughness

We consider here the class of sounds which can be modeled by a sum of sinusoids with fixed frequencies. We then exclude sounds containing a stochastic part, for which other methods of roughness computation should be used [10]. Roughness is computed from the amplitudes and frequencies values of spectral components contained in the sound (cf. Eq. 2). Since different sets of synthesis parameters (amplitudes, frequencies) can produce sounds with the same roughness, controlling the roughness is not obvious. Moreover, as the damping parameters α_G and α_R act on the amplitudes, they also act on the roughness, which means that damping and roughness cannot be totally decorrelated.

In our synthesis model implementation, the generated spectrum is composed of 64 components and R depends on 128 parameters (64 values of amplitudes and frequencies). To investigate the behavior of R as function of the characteristics of the spectrum, we reduced the number of parameters by considering specific laws for amplitudes and frequencies. We constrained the set of frequencies to a fundamental one, f_1, and partials following an inharmonicity law driven by a single γ parameter. The frequencies f_k are obtained by : $f_k = f_1 k^\gamma$ where $k = 1, \ldots, 64$. While $\gamma = 1$

gives a perfect harmonic series, the variation of γ allow to stretch the distribution of partials along the frequency axis. This different values of γ allow for the generation of a large panel of sounds, from plates (γ < 1) to strings ($\gamma \approx$ 1) or bells (γ >1). Similarly, the amplitudes A_k (expressed in dB) are defined by $A_k = A_1 + \theta(k-1)$, where A_1 is the amplitude of the fundamental component and θ the slope characterizing the envelope spectrum ($\theta < 0$ generates a “low-pass" spectrum and $\theta > 0$ a “high-pass" spectrum). With these considerations, we examined the evolution of R as a function of the reduced set of synthesis parameters $\{f_1, A_1, \gamma, \theta\}$. Figure 2 **(a)** shows the evolution of R as a function of γ (from 0.1 to 2) and θ (from -1 to 1) with f_1 fixed to 27.5 Hz (corresponding to the note MIDI A-1). Figure 2 **(b)** shows the evolution of R as a function of γ (from 0.1 to 2) and f_1 (from note A-1 (27.5 Hz) to A6 (3520 Hz)).

Obviously, a given roughness can be obtained with many combinations of γ and θ, hence, this function is not invertible. It also appears that R is maximum for allpass type spectra ($\theta = 0$). Indeed, it can be shown from Equ. Eq. 2 that the roughness of a pair of sines is maximized when the amplitudes are of the same magnitude order, and is higher for higher amplitudes. Applying a lowpass or highpass filter on a flat spectrum will then reduce contributions of high and low partials respectively, hence leading to a smaller value of R.

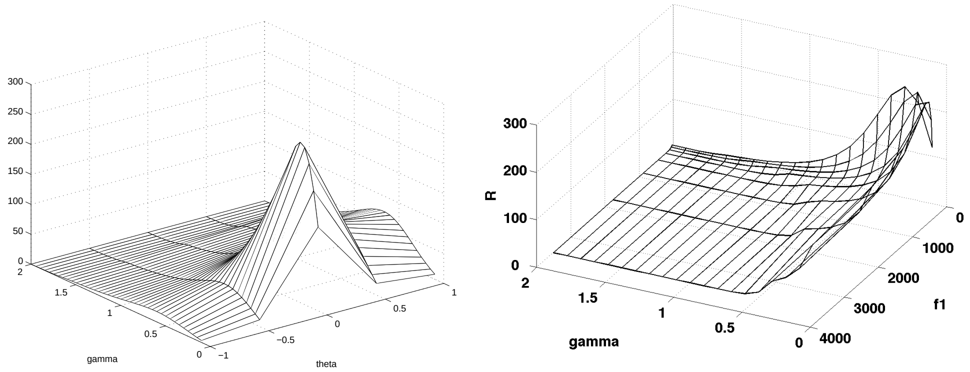

Figure 2: **(a)** Roughness as a function of the envelope slope parameter θ (from -1 to 1) and the inharmonicity parameter γ (from 0.1 to 2). f_1 is fixed at 27.5 Hz (A-1). **(b)** Roughness as a function of γ (from 0.1 to 2) and f_1 ranging from note A-1 (27.5 Hz) to A6 (3520 Hz).

By studying the evolution of R as a function of the fundamental frequency f_1, we noticed that when f_1 is increased, the whole spectrum is translated towards high frequencies making higher partials leaving the audio bandwith, hence reducing R. This can be seen on figure 2 **(b)** where R is represented as a function of γ and f_1. Except the small region where both f_1 and γ are low, R is a monotonic function of γ for a fixed value of f_1, leading to a real time control of R.

CONCLUSIONS

We here proposed a mapping strategy for the control of the "material element" of a synthesis model dedicated to impact sounds. We aimed at offering an intuitive manipulation of synthesis parameters in terms of material "coloration" (wooden, metallic or crystal-clear sounds). For that, we designed a sound categorization test to determine sounds judged as typical of different material categories such as Wood, Metal or Glass. The acoustic analysis of these typical sounds revealed the importance of several timbre descriptors: the damping parameters and the roughness for the categorization of different materials. The calibration of these timbre descriptors from behavioral data allowed to define specific domains for the generation of ecological sounds evoking different materials. The control of the “material element" was effectuated by the manipulation of these timbre descriptors. In future studies, we aim at investigating mapping strategies dedicated to the “object element" in terms of its perceived size or shape. Actually, in a musical context, we have proposed to control the synthesis parameters corresponding to the oscillator banks by using a chord generator [3].

Acknowledgement

This projects has partly been supported by the French National Research Agency (ANR, JC05-41996, "senSons") to Sølvi Ystad.

References

[1] M. Aramaki, M. Besson, R. Kronland-Martinet, S. Ystad. Timbre perception of sounds from impacted materials: a neuro-acoustic approach. Journal of the Acoustical Society of America **submitted (2007)**

[2] M. Aramaki, R. Kronland-Martinet. Analysis-synthesis of impact sounds by real-time dynamic filtering. IEEE Transactions on Speech and Audio Processing **14 (2006), no. 2**

[3] M. Aramaki, R. Kronland-Martinet, T. Voinier, S. Ystad. A percussive sound synthetizer based on physical and perceptual attributes. Computer Music Journal **30 (2006), no. 2** 32–41

[4] A. Caracciolo, C. Valette. Damping mechanisms governing plate vibration. Acta Acustica **3 (1995)** 393–404

[5] C. Carello, K. L. Anderson, A. J. Kunkler-Peck. Perception of object length by sound. Psychological Science **9 (1998), no. 3** 211–214

[6] D. J. Freed. Auditory correlates of perceived mallet hardness for a set of recorded percussive events. Journal of the Acoustical Society America **87 (1990), no. 1** 311 – 322

[7] B. L. Giordano, S. McAdams. Material identification of real impact sounds: Effects of size variation in steel, wood, and plexiglass plates. Journal of the Acoustical Society of America **119 (2006), no. 2** 1171–1181

[8] R. L. Klatzky, D. K. Pai, E. P. Krotkov. Perception of material from contact sounds. Presence: Teleoperators and Virtual Environments **9 (2000), no. 4** 399–410

[9] A. J. Kunkler-Peck, M. T. Turvey. Hearing shape. J. of Experimental Psychology: Human Perception and Performance **26 (2000), no. 1** 279–294

[10] M. Leman. Visualization and calculation of the roughness of acoustical musical signals using the synchronisation index model (sim). In *Proceedings of the COST-G6 Conference on Digital Audio Effects (DAFX-00)* (2000)

[11] S. McAdams, S. Winsberg, S. Donnadieu, G. D. Soete, J. Krimphoff. Perceptual scaling of synthesized musical timbres: common dimensions, specificities, and latent subject classes. Psychological Research **58 (1995)** 177–192

[12] D. Rocchesso, F. Fontana. *The Sounding Object*. D. Rocchesso and F. Fontana (2003)

[13] J. O. Smith, S. A. van Duyne. Developments for the commuted piano. In *Proceedings of the International Computer Music Conference* (1995)

[14] K. van den Doel, D. K. Pai. The sounds of physical shapes. Presence **7 (1998), no. 4** 382 – 395

[15] P. N. Vassilakis. *Perceptual and Physical Properties of Amplitude Fluctuation and their Musical Significance*. Ph.D. thesis, University of California, Los Angeles (2001)

[16] D. L. Wessel. Timbre space as a musical control structure. Computer Music Journal **3 (1979), no. 2** 45–52

[17] R. P. Wildes, W. A. Richards. *Recovering material properties from sound*. W. A. Richards Ed., MIT Press, Cambridge (1988) 356–363, 356–363

[18] E. Zwicker, H. Fastl. *Psychoacoustics, facts and models*. Springer-Verlag (1990)